Bild- und Toncodierung für die Multimedia-Kommunikation

Musmann, Hans Georg

Veröffentlicht in:
Jahrbuch 2001 der Braunschweigischen Wissenschaftlichen Gesellschaft, S.95-103

J. Cramer Verlag, Braunschweig

HANS GEORG MUSMANN, Hannover*

Bild- und Toncodierung für die Multimedia-Kommunikation

1. Einleitung

Das weltweite Internet ermöglicht in Zukunft nicht nur die Übertragung von Sprache, Texten und Textbildern, sondern auch von Bewegtbildern, Musik und sogar dreidimensionalen Informationen virtueller Objekte. Die elektrischen analogen Signale dieser Multimedia-Information müssen dazu in eine digitale Darstellung gewandelt werden.

Bild 1 veranschaulicht den Vorgang der Digitalisierung eines analogen zeitveränderlichen Signals. Das analoge Signale einer Nachricht wird dazu in äquidistanten Zeitabschnitten abgetastet. Die Amplitude eines jeden Abtastwertes entspricht der Lautstärke bei einem Sprachsignal bzw. Helligkeit eines Bildpunktes beim Fernsehsignal. Die Amplitude wird in Stufen quantisiert und jeder Stufe eine Dualzahl zugeordnet. Die Ziffern der Dualzahlen werden durch Binärsymbole 1 und 0 codiert. Verschiedene Nachrichtensignale können auf diese Weise in eine Folge von Binärsymbolen (bit) umgewandelt und aus dieser Folge auch wieder zurückgewonnen werden. Diese digitale Darstellungsform einer Nachricht wird als Pulscodemodulation (PCM) bezeichnet.

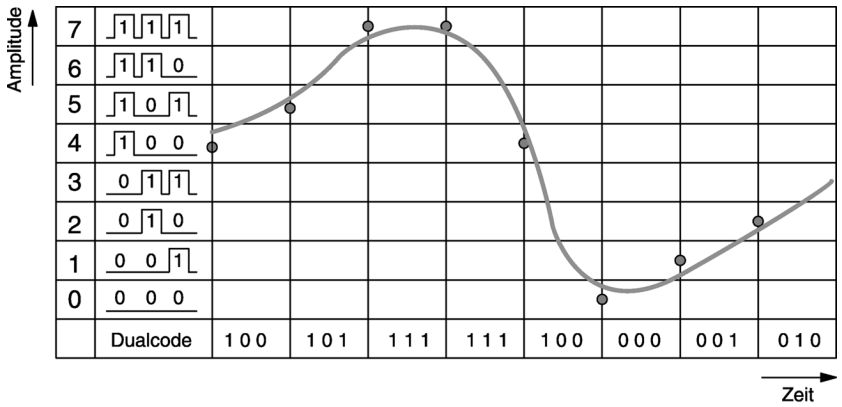

Bild 1: Quantisierung und Codierung von Signalen mit PCM

* Vortrag gehalten beim Kolloquium anläßlich der Jahresversammlung der Braunschweigischen Wissenschaftlichen Gesellschaft am 18. Mai 2001.

Der besondere Vorzug der digitalen Darstellungsform liegt zum einen in der größeren erreichbaren Genauigkeit der Signaldarstellung und der damit verbundenen höheren Ton- und Bildqualität, wie sie beispielsweise von der Compact Disc her bekannt ist. Zum anderen erlaubt die binäre Darstellungsform, dass in Zukunft alle Nachrichtensignale über ein und dasselbe Nachrichtennetz übertragen und ein derartiges Netz somit für die Multimedia-Kommunikation verwendet werden kann.

Ein Problem bilden dabei Bewegtbildsignale wie das Fernsehsignal, da sie eine relativ große Übertragungsbitrate benötigen.Vergleichsweise erfordert die Übertragung eines Videosignals in TV-Auflösung mit 166 Mbit/s die Übertragungsrate von etwa 2500 Fernsprechsignalen von je 64 kbit/s und damit entsprechend hohe Übertragungskosten.

Für eine breite Anwendung der Multimedia-Kommunikation mussten daher zunächst effiziente Verfahren der datenreduzierenden Bild- und Toncodierung entwickelt werden. In den vergangenen Jahren wurden die ersten Codierungsverfahren von der International Standardization Organization (ISO) standardisiert.

Nachfolgend werden die Konzepte der standardisierten Audio- und Videocodierungen und die daraus hervorgegangenen neuen Kommunikationsdienste und Anwendungen kurz beschrieben. Abschließend wird ein Ausblick auf die laufenden Forschungsarbeiten und deren Anwendungen gegeben.

2. Audio-Codierung

Bild 2 zeigt die PCM-Formate einiger Ton- und Sprachsignale

Die digitale Darstellung eines stereophonen Audiosignals im Studioformat erfordert eine Abtastfrequenz von 48 kHz und eine gleichförmige Quantisierung entsprechend 16 bit pro Abtastwert [1]. Daraus resultiert eine Datenrate von 768 kbit/s für ein Monosignal und entsprechend 2 × 768 kbit/s, also etwa 1,5 Mbit/s für ein Stereosignal. Aufgrund der etwas

PCM–FORMAT	DVD–AUDIO (TYP.)	DAT (TYP.)	CD	FERNSPRECHEN (7KHZ)	FERNSPRECHEN ITU–T G.711
ABTAST–FREQUENZ	96 KHZ	48 KHZ	44,1 KHZ	16 KHZ	8 KHZ
QUANTISIERUNG DER ABTASTWERTE	24 BIT GLEICHFÖRMIG	16 BIT GLEICHFÖRMIG	16 BIT GLEICHFÖRMIG	16 BIT GLEICHFÖRMIG	8 BIT UNGLEICH–FÖRMIG
KANALANZAHL	2	2	2	1	1
BITRATE	2·2,3 MBIT/S	2·768 KBIT/S	2·706 KBIT/S	256 KBIT/S	64 KBIT/S

Bild 2: PCM-Formate für Ton- und Sprachsignale

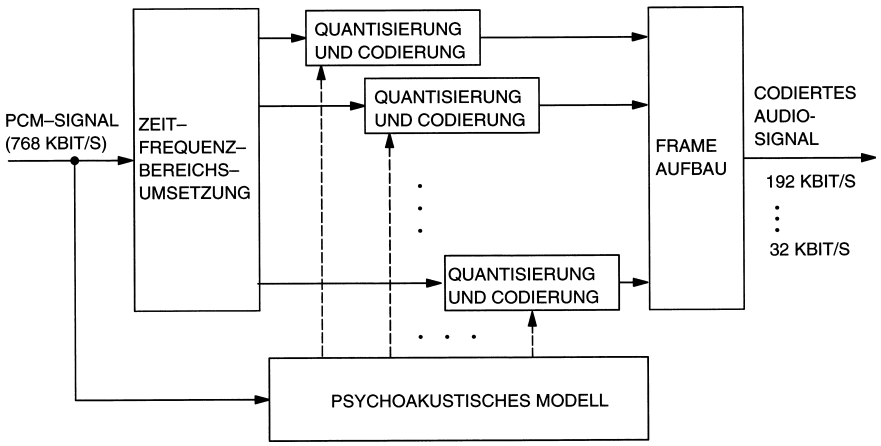

Bild 3: Generelles Blockschaltbild des Audio Coders

geringeren Abtastfrequenz von 44,1 kHz ergibt sich für das Stereosignal einer Compact Disc eine Datenrate von 2 × 706 kbit/s, also etwa 1,4 Mbit/s.

Der von der ISO entwickelte Audio-Codierungsstandard zur Reduktion der Datenrate besteht aus drei Layern, wobei Komplexität und Codierungseffizienz von Layer I zu Layer II und Layer III jeweils zunimmt [2]. Hier soll nur das grundlegende Codierungsprinzip erläutert werden.

Bild 3 zeigt das allgemeine Blockschaltbild eines Audio-Coders. Zunächst wird das am Eingang eingespeiste PCM-Audiosignal aus dem Zeitbereich in eine Darstellung im Frequenzbereich umgesetzt, was einer Zerlegung des Eingangssignals in Spektralkomponenten entspricht. Eine ähnliche Zerlegung in die sog. Frequenzgruppen wird auch vom menschlichen Gehör bei der Wahrnehmung von Tonsignalen vorgenommen. Anschließend wird jede aus der Zeit-Frequenzbereichs-Umsetzung hervorgegangene Spektralkomponente individuell entsprechend den aktuellen Maskierungseigenschaften des Gehörs quantisiert und codiert. Die Steuerinformation für die Quantisierer wird dabei aus einer Schätzung der aktuellen signalabhängigen Maskierungsschwelle gewonnen, die parallel zur Zeit-Frequenzbereichs-Umsetzung vom psychoakustischen Modell berechnet wird. Die Maskierungsschwelle gibt die maximal erlaubte Quantisierungsfehlerleistung für jede einzelne Spektralkomponente an. Die Quantisierung ist optimal, wenn jede Spektralkomponente exakt mit der durch die Maskierungsschwelle vorgegebenen Genauigkeit quantisiert und codiert wird. In diesem Fall wird das dem Nutzsignal überlagerte Quantisierungsgeräusch vom menschlichen Gehör gerade noch nicht wahrgenommen.

Die Zeit-Frequenzbereichs-Umsetzung kann entweder mit einer Filterbank oder einer Transformation oder auch einer Kombination von beiden realisiert werden. Zur Gewährleistung einer effizienten Codierung ist es erforderlich, dass die Anzahl der zu codierenden Spektralkomponenten in jedem Zeitintervall nicht größer ist als die Anzahl der Abtastwerte

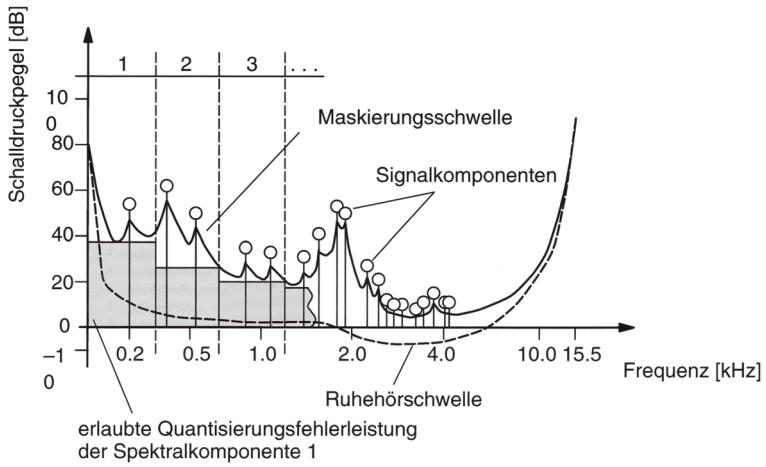

Bild 4: Maskierungsschwelle für den Vokal „A“

des Eingangssignals. Durch den Einsatz von Filterbänken mit Abtastratenreduktion und Aliasing-Kompensation kann dieses Problem gelöst werden.

Das psychoakustische Modell berechnet jeweils für kurze Zeitintervalle des PCM-Eingangssignals einen Schätzwert für die aktuelle signalabhängige Maskierungsschwelle. Für jede Spektralkomponente gibt die Maskierungsschwelle im zugehörigen Frequenzintervall die gerade noch wahrnehmbare Quantisierungsfehlerleistung an. Diese wird dann zur Generierung der Steuerinformation für die dynamische Bit- bzw. Rauschleistungszuweisung verwendet, indem die aktuell erforderliche Anzahl der Quantisierungsstufen für jede Spektralkomponente und jedes Zeitintervall bestimmt wird.

Ein Beispiel für die Berechnung einer Maskierungsschwelle eines Zeitintervalls zeigt Bild 4. Dargestellt sind die Ruhehörschwelle, die Signalkomponenten des Vokals „A“ und die zugehörige globale Maskierungsschwelle. Die grau hinterlegten Bereiche geben jeweils die durch eine geeignete Bitzuweisung eingestellte Quantisierungsfehlerleistung für die ersten Spektralkomponenten an.

Mit dem sog. MPEG 1 Layer III Codierungsstandard konnte die Datenrate eines stereophonen Audiosignals von 1,5 Mbit/s auf 2 × 128 kbit/s = 256 kbit/s reduziert werden und dabei eine der Compact Disc vergleichbare Audioqualität gewährleistet werden. Erst der nachfolgende Codierungsstandard MPEG 2 AAC erreichte die angestrebte Datenrate von 2 × 64 kbit/s = 128 kbit/s [3].

Aus diesen Fortschritten der Audiocodierung ergaben sich im wesentlichen 3 neue Anwendungen.

1. Digitale Audiosignale hoher Tonqualität können im Selbstwähldienst über einen ISDN Basisanschluss übertragen werden.

2. Digitale Audiosignale hoher Tonqualität können mit nur geringen Verzögerungen im Internet übertragen werden.
3. Digitale Audiosignale hoher Tonqualität können in einem Audio-Recorder ohne Laser und Motor, ausgerüstet mit einem Speicherchip, aufgezeichnet und wiedergegeben werden.

3. Video-Codierung

Bild 5 zeigt die PCM-Formate von Video-Signalen unterschiedlicher Auflösungen.

Die digitale Darstellung eines Studio-TV-Signals erfordert gemäß der CCIR-Recommendation 601 [4] eine Nettodatenrate von 166 Mbit/s. Zur Übertragung von Bewegtbildern für Bildtelefon- und Videokonferenzanwendungen mit weit geringeren Datenraten wurden das Common Intermediate Format CIF und QCIF eingeführt.

Der CCITT-Codierungsstandard H.261 beschreibt ein Codierungsverfahren mit dem Bildfernsprechsignale im Common Intermediate Format (CIF) bei reduzierter Bildfolgefrequenz von 10 Hz mit einer Datenrate von $p \times 64$ kbit/s übertragen werden können. Für den Fall $p = 1$ kann damit ein B-Kanal eines ISDN-Basisanschlusses für die Bildübertragung genutzt werden, während der zweite B-Kanal für die Sprachübertragung zur Verfügung steht.

Erstes Ziel der ISO Motion Picture Expert Group (MPEG) war die Codierung von Bewegtbildern mit Begleitton für digitale Speichermedien bei einer Datenrate von bis zu 1,5 Mbit/s ebenfalls unter Verwendung von CIF.

Den beiden genannten Video-Codierungsverfahren liegt das gleiche Konzept der bewegungskompensierten Hybridcodierung zugrunde. Es besteht im wesentlichen aus zwei Komponenten. Mit einer bewegungskompensierten Prädiktion wird das nächste zu über-

	CCIR 601 – 625	CIF–625	CIF	QCIF
ANZAHL DER ABTAST-WERTE JE ZEILE				
LUMINANZ:	720	352	352	176
CHROMINANZ:	360	176	176	88
ANZAHL DER AKTIVEN ZEILEN JE BILD				
LUMINANZ:	288	288	288	144
CHROMINANZ:	288	144	144	72
QUANTISIERUNG IN BIT JE ABTASTWERT:	8	8	8	8
BILDFOLGEFREQUENZ:	50 ZEILEN–VERSCHRÄNKT	25 NICHT ZEILEN-VERSCHRÄNKT	10 NICHT ZEILEN-VERSCHRÄNKT	10 NICHT ZEILEN-VERSCHRÄNKT
BITRATE IN MBIT/S:	166	30.5	12.1	3.0

Bild 5: PCM-Formate für Videosignale

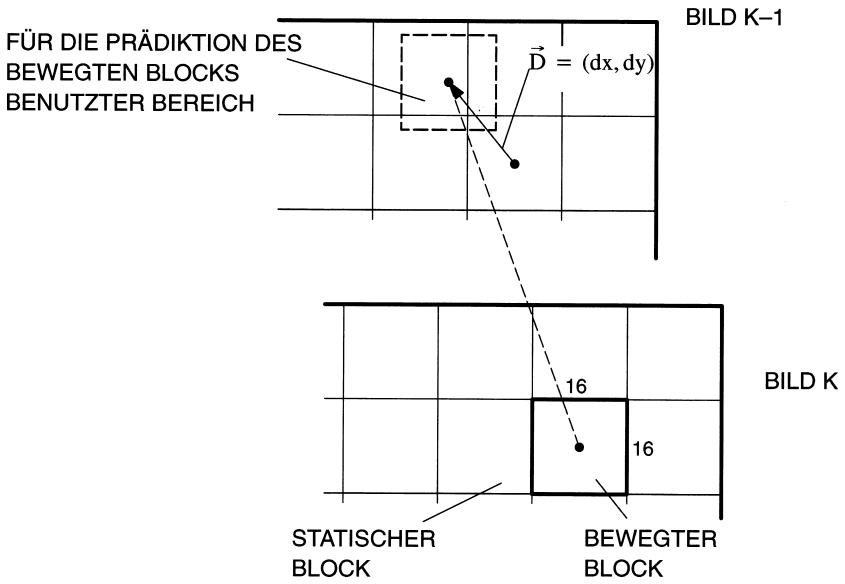

Bild 6: Illustration einer bewegungskompensierenden Prädiktion

tragende Bild k aus dem vorangegangenem und bereits übertragenen Bild k-1 vorhergesagt. Zu diesem Zweck wird das zu übertragende Bild k in Blöcke von 16 × 16 Bildpunkten unterteilt, siehe Bild 6.

Für jeden Block wird auf der Senderseite ein sog. Displacementvektor $\bar{D}$ gemessen und zum Empfänger übertragen. Der Displacementvektor zeigt an, wo die Bildinformation eines Blockes im vorangegangenen Bild war. Mit Hilfe des Displacementvektors und dem vorangegangenen Bild k-1 kann nun im Coder und im Decoder ein Vorhersagebild konstruiert werden, indem die Bildinformation blockweise aus dem vorangegangenen Bild k-1 entsprechend der Displacementvektoren übernommen wird. Im Coder kann das Vorhersagebild mit dem zu übertragenden Bild k durch Differenzbildung verglichen werden. Bild 7 zeigt als Beispiel in Gegenüberstellung das zu übertragende Bild k und das Differenzbild, das sog. Prädiktionsfehlerbild. Nur das Prädiktionsfehlerbild muss zum Empfänger übertragen werden. Durch Addition des Prädiktionsfehlerbildes zum Vorhersagebild kann der Decoder das Empfangsbild k rekonstruieren.

Um die Übertragung des Prädiktionsfehlerbildes mit geringer Datenrate auszuführen, wird das Prädiktionsfehlerbild in Blöcke von 8 × 8 Bildpunkten unterteilt und jeder Block einer Discreten Cosinus Transformation (DCT) unterzogen, siehe Bild 8. Übertragen werden die sich ergebenden 8 × 8 DCT-Koeffizienten. Die DCT bewirkt, dass viele der DCT-Koeffizienten Null sind und dafür nur eine geringe Datenrate erforderlich ist.

Diese bewegungskompensierte Hybridcodierung wird auch im ISO MPEG-2 Codierungsstandard zur Codierung von Videosignalen in Fernsehauflösung gemäß

Bild 7: Eingangsbild und Prädiktionsfehlerbild

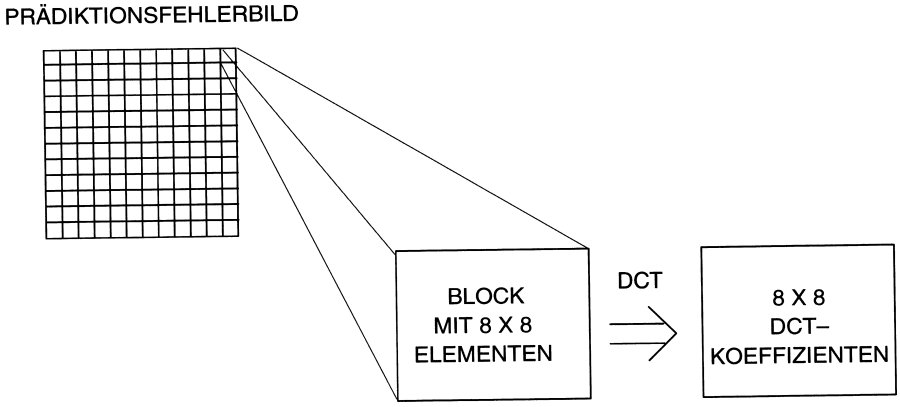

Bild 8: Diskrete Cosinus Transformation (DCT) der Prädiktionsfehler

CCCIR 601 angewendet [5]. Als Ergebnis konnten die Daten von 166 Mbit/s auf 4 Mbit/s reduziert werden und dabei die Bildqualität des Fernsehrundfunks bewahrt werden.

Aus diesen Fortschritten der Videocodierung ergaben sich im wesentlichen folgende neue Dienste und Anwendungen

1. Ein analoger Fernsehkanal kann 6 digitale Fernsehsignale übertragen.
2. Die Einführung des digitalen Fernsehrundfunks.
3. Eine Compact Disc kann Audio- und Videosignale aufzeichnen, bekannt als DVD.
4. Ein Computer kann Videosignale speichern.

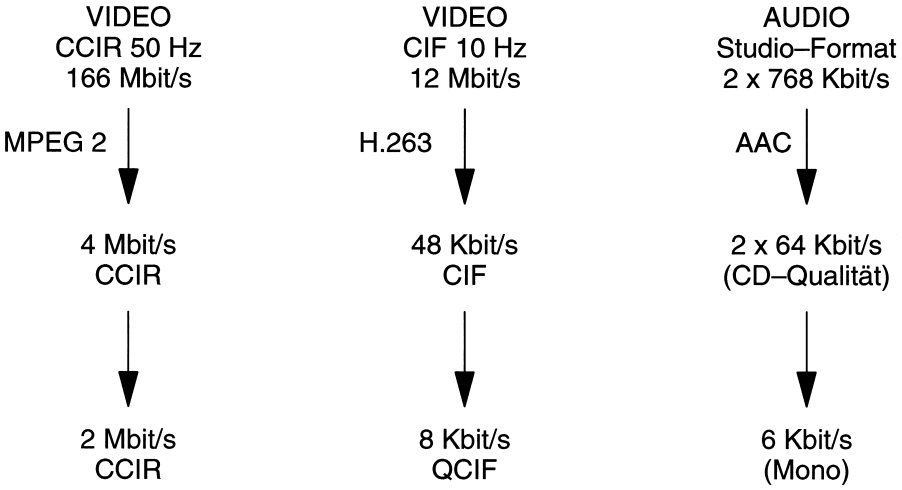

Bild 9: Ziele der derzeitigen Forschung

4. Zusammenfassung und Ausblick

Bild 9 zeigt in einem Überblick die erreichten Ergebnisse und die Ziele der derzeitigen Forschung.

Die Reduzierung der Datenrate eines Videosignals nach CCIR 601 auf 2 Mbit/s ist inzwischen schon erreicht. In [6] wird sogar nachgewiesen, dass bezogen auf MPEG-4, eine Weiterentwicklung von MPEG-2, ein zusätzlicher Reduktionsfaktor von 2 erreicht worden ist. Sobald die Datenrate in die Größenordnung der Datenrate eines DSL-Anschlusses kommt, können Fernsehsignale aus dem Internet über die Teilnehmeranschlussleitungen des Fernsprechnetzes zum Teilnehmer übertragen werden. Das Fernsehen wird damit ganz neue Möglichkeiten erhalten.

Sobald ein Videosignal in QCIF-Darstellung mit 8 kbit/s und ein monophones Audiosignal mit 6 kbit/s in ausreichender Qualität codiert werden können, besteht die Möglichkeit Fernsehsignale in kleiner Bildgröße auch über das Mobilfunknetz zu übertragen und auf einem Handy darzustellen. Aufgrund der kleinen Bildgröße wird ein derartiger Dienst nicht so sehr für die Übertragung von Unterhaltungssendungen, aber für Nachrichtensendungen von Interesse sein, da sie weltweit empfangen werden können.

Literaturverzeichnis

[1] CCIR, Source encoding for digital sound signals in broadcasting studios, CCIR Recommendation 646, Geneva 1986

[2] MPEG-1, Coding of moving pictures and associated audio for digital storage media at up to 1,5 Mbit/s, part 3: Audio . International Standard IS 11172-3, ISO/IEC JTC/SC 29 WG 11, 1992

[3] MPEG-2, Advanced audio coding, AAC International Standard IS 13818-7, ISO/IEC JTC/SC 29 WG 11, 1997

[4] CCIR, Encoding parameters of digital television for studios, CCIR Recommendation 601, Geneva 1982

[5] MPEG-2, Generic Coding of moving pictures and associated audio, Part 2 Video, International Standard IS 18818, ISO/IEC JTC/SC 29 WG 11, 1998

[6] MPEG-4, Summary Information for ITU-T VCEG Draft H.26L Algorithm in Response to Video and DCinema CfPs, Doc MPEG 2001/M7511, ISO/IEC JTC1/SC 29 WG11, July 2001

Prof. Dr.-Ing. H. G. Musmann
Institut für Theoretische Nachrichtentechnik
und Informationsverarbeitung · Universität Hannover
Appelstraße 9 A · D-30167 Hannover